## B. 効果ベクトル $\boldsymbol{\beta}$, デザイン行列 $\boldsymbol{X}$ の構成

ここでは，変数 $A, B, C, D$ よりつくられる四元分割表を扱う．二元分割表や三元分割表には，以下の説明のなかからそれぞれ変数 $A, B$ または変数 $A, B, C$ に関する部分を選択適用すればよい．五元もしくはそれ以上の高次の分割表への拡張もそれほど難しいことではないので，必要に応じて試してもらいたい．対数線型モデルをデータの実証分析に用いるには，SPSS-X(LOGLINEAR) や SAS(CATMOD) などの汎用パッケージが便利である．これらのパッケージでは利用者がデザイン行列 $\boldsymbol{X}$ を設定できるように配慮されていて，柔軟な分析が可能である．こうした機能を享受するには，やはりデザイン行列の構成原理を知る必要がある．たとえその機能を活用しなくとも，通し番号で出力される効果パラメータの情報を解釈する上で，$\boldsymbol{X}$ と対応する $\boldsymbol{\beta}$ の要素がどう並んでいるのか理解しておくと役に立つであろう．

変数 $A, B, C, D$ はそれぞれ $I, J, K, L$ 個のカテゴリーをもつものとする．これらのカテゴリーを添字 $i, j, k, l$ で指示する：

$$i=1, \cdots, I; \quad j=1, \cdots, J; \quad k=1, \cdots, K; \quad l=1, \cdots, L$$

カテゴリー集合をそれぞれ $\mathscr{I}, \mathscr{J}, \mathscr{K}, \mathscr{L}$ で表すとセルの集合は

$$\{(i, j, k, l) : i \in \mathscr{I},\ j \in \mathscr{J},\ k \in \mathscr{K},\ l \in \mathscr{L}\}$$

と書ける．以下，最も包括的な飽和モデルについて説明する．不飽和モデルへの応用は特に難しいものではない．

**1. ダミーコーディング：** 基準セル$=(i^*, j^*, k^*, l^*)$

四元分割表の飽和モデル $[ABCD]$

$$\begin{aligned}\log m_{ijkl} = {} & u + u_i^A + u_j^B + u_k^C + u_l^D + u_{ij}^{AB} + u_{ik}^{AC} + u_{il}^{AD} + u_{jk}^{BC} + u_{jl}^{BD} + u_{kl}^{CD} \\ & + u_{ijk}^{ABC} + u_{ijl}^{ABD} + u_{ikl}^{ACD} + u_{jkl}^{BCD} + u_{ijkl}^{ABCD} \\ & (i \in \mathscr{I},\ j \in \mathscr{J},\ k \in \mathscr{K},\ l \in \mathscr{L}) \end{aligned} \tag{B.1}$$

である．ここで，基準カテゴリーを除いた別の添字の集合 $\underline{\mathscr{I}}, \underline{\mathscr{J}}, \underline{\mathscr{K}}, \underline{\mathscr{L}}$ を定義する．

$$\underline{\mathscr{I}} = \{1, \cdots, I\} - \{i^*\}, \qquad \underline{\mathscr{J}} = \{1, \cdots, J\} - \{j^*\}$$
$$\underline{\mathscr{K}} = \{1, \cdots, K\} - \{k^*\}, \qquad \underline{\mathscr{L}} = \{1, \cdots, L\} - \{l^*\}$$

見込み(odds)に基づく効果項の定義より，

$$i=i^* \text{ or } j=j^* \text{ or } k=k^* \text{ or } l=l^* \quad \text{のとき}$$
$$u_i^A = u_j^B = u_k^C = u_l^D = 0, \quad u_{ij}^{AB} = u_{ik}^{AC} = u_{il}^{AD} = u_{jk}^{BC} = u_{jl}^{BD} = u_{kl}^{CD} = 0$$
$$u_{ijk}^{ABC} = u_{ijl}^{ABD} = u_{ikl}^{ACD} = u_{jkl}^{BCD} = u_{ijkl}^{ABCD} = 0$$

なので，独立な効果項の添字については

$$i \in \underline{\mathscr{I}}, \quad j \in \underline{\mathscr{J}}, \quad k \in \underline{\mathscr{K}}, \quad l \in \underline{\mathscr{L}}$$

が適用される．

ちなみに，$u_{ij}^{AB}$ の添字 $ij$ は $\langle i, j \rangle$ と記すのが作法に則した書き方で，変化の範囲は $\{\langle i, j \rangle : i \in \underline{\mathscr{I}},\ j \in \underline{\mathscr{J}}\}$ となる．$u_{ijk}^{ABC}$, $u_{ijkl}^{ABCD}$ の添字についても同じく

$$\{\langle i, j, k \rangle : i \in \underline{\mathscr{I}}, j \in \underline{\mathscr{J}}, k \in \underline{\mathscr{K}}\}, \qquad \{\langle i, j, k, l \rangle : i \in \underline{\mathscr{I}}, j \in \underline{\mathscr{J}}, k \in \underline{\mathscr{K}}, l \in \underline{\mathscr{L}}\}$$

を用いるのが正しい．このように集合の要素を掛け合わせたものを要素にしてできる集合を直積とよび

$$\underline{\mathscr{I}}\times\underline{\mathscr{J}}=\{\langle i,j\rangle : i\in\underline{\mathscr{I}},\ j\in\underline{\mathscr{J}}\}$$
$$\underline{\mathscr{I}}\times\underline{\mathscr{J}}\times\underline{\mathscr{K}}=\{\langle i,j,k\rangle : i\in\underline{\mathscr{I}},\ j\in\underline{\mathscr{J}},\ k\in\underline{\mathscr{K}}\}$$
$$\underline{\mathscr{I}}\times\underline{\mathscr{J}}\times\underline{\mathscr{K}}\times\underline{\mathscr{L}}=\{\langle i,j,k,l\rangle : i\in\underline{\mathscr{I}},\ j\in\underline{\mathscr{J}},\ k\in\underline{\mathscr{K}},\ l\in\underline{\mathscr{L}}\}$$

と表現する．他の添字についても同様である．このことを踏まえた上で，以下の説明では意味を損なわない限り，表現を簡略化する：たとえば，$ijk=\langle i,j,k\rangle$．また，このように組み合わせられた複数の添字の変化の優先順位は，本文中の記述と統一するために

$$j\Rightarrow i\Rightarrow k\Rightarrow l$$

とする．すなわち，$i^*=1$，$j^*=1$，$k^*=1$ のとき

$$\underline{\mathscr{I}}\times\underline{\mathscr{J}}\times\underline{\mathscr{K}}=\{222, 232, \cdots, 2J2, 322, \cdots, 3J2, \cdots, IJ2, \cdots, IJK\}$$

## 1.1　効果ベクトル $\beta$

飽和モデル (B.1) の効果項を関与する変数の数に応じて次のように分類する：

$$u^0=\{u\},$$
$$u^1=\{u_i^A, u_j^B, u_k^C, u_l^D : i\in\underline{\mathscr{I}},\ j\in\underline{\mathscr{J}},\ k\in\underline{\mathscr{K}},\ l\in\underline{\mathscr{L}}\}$$
$$u^2=\{u_{ij}^{AB}, u_{ik}^{AC}, u_{il}^{AD}, u_{jk}^{BC}, u_{jl}^{BD}, u_{kl}^{CD} : i\in\underline{\mathscr{I}},\ j\in\underline{\mathscr{J}},\ k\in\underline{\mathscr{K}},\ l\in\underline{\mathscr{L}}\}$$
$$u^3=\{u_{ijk}^{ABC}, u_{ijl}^{ABD}, u_{ikl}^{ACD}, u_{jkl}^{BCD} : i\in\underline{\mathscr{I}},\ J\in\underline{\mathscr{J}},\ k\in\underline{\mathscr{K}},\ l\in\underline{\mathscr{L}}\}$$
$$u^4=\{u_{ijkl}^{ABCD} : i\in\underline{\mathscr{I}},\ j\in\underline{\mathscr{J}},\ k\in\underline{\mathscr{K}},\ l\in\underline{\mathscr{L}}\}$$

集合の要素の順序は本来意味をもたないが，効果ベクトルを形成するときの都合を考え，一応ここでは簡単に上の各集合について，並べられた主効果，交互作用ごとに添字の範囲を尽くすことにする．たとえば，かりに基準カテゴリーが (1, 1, 1, 1) なら

$$u^1=\{u_2^A, \cdots, u_I^A, u_2^B, \cdots, u_J^B, u_2^C, \cdots, u_K^C, u_2^D, \cdots u_L^D\}$$

を意味する．

効果項の集合の要素数を求めておこう．

$$u^0=1, \qquad u^1=(I-1)+(J-1)+(K-1)+(L-1)$$
$$u^2=(I-1)(J-1)+(I-1)(K-1)+(I-1)(L-1)+(J-1)(K-1)+(J-1)(L-1)+(K-1)(L-1)$$
$$u^3=(I-1)(J-1)(K-1)+(I-1)(J-1)(L-1)+(I-1)(K-1)(L-1)+(J-1)(K-1)(L-1)$$
$$u^4=(I-1)(J-1)(K-1)(L-1)$$

当然，$u^0$ から $u^4$ までの和集合 $u$ はすべての効果項の和からなり

$$u=u^0\cup u^1\cup u^2\cup u^3\cup u^4$$

その要素数は $IJKL$ である．

効果ベクトル $\beta$ もその成分を効果の次数で小行列 $\beta^0, \beta^1, \beta^2, \beta^3, \beta^4$ に分割する：

$$\boldsymbol{\beta}=\begin{bmatrix}\boldsymbol{\beta}^0\\ \boldsymbol{\beta}^1\\ \boldsymbol{\beta}^2\\ \boldsymbol{\beta}^3\\ \boldsymbol{\beta}^4\end{bmatrix}$$

$\boldsymbol{\beta}^0=[u]$ は明らかである．$\boldsymbol{\beta}^1$ は効果項の集合 $\boldsymbol{u}^1$ の要素を先頭から順に成分とする

$$[(I-1)+(J-1)+(K-1)+(L-1)]\times 1$$

の行列である．$\boldsymbol{\beta}^2, \boldsymbol{\beta}^3, \boldsymbol{\beta}^4$ と $\boldsymbol{u}^2, \boldsymbol{u}^3, \boldsymbol{u}^4$ との対応も同様である．

## 1.2 デザイン行列 $\boldsymbol{X}$

デザイン行列は期待頻度（の対数）

$$\boldsymbol{\theta}=\begin{bmatrix}\log m_{1111}\\ \log m_{1211}\\ \vdots\\ \log m_{IJKL}\end{bmatrix},\qquad \text{または，}\ \boldsymbol{\theta}=[\log m_s]\ \ (s\in\underset{\#}{\mathscr{I}}\times\underset{\#}{\mathscr{J}}\times\underset{\#}{\mathscr{K}}\times\underset{\#}{\mathscr{L}})$$

を効果ベクトルに結びつける役割をする．行列 $\boldsymbol{X}$ は各効果項に対応して細分化するとわかりやすくなる．

$$\boldsymbol{X}=$$
$$[\boldsymbol{X}^0|\boldsymbol{X}^A|\boldsymbol{X}^B|\boldsymbol{X}^C|\boldsymbol{X}^D|\boldsymbol{X}^{AB}|\boldsymbol{X}^{AC}|\boldsymbol{X}^{AD}|\boldsymbol{X}^{BC}|\boldsymbol{X}^{BD}|\boldsymbol{X}^{CD}|\boldsymbol{X}^{ABC}|\boldsymbol{X}^{ABD}|\boldsymbol{X}^{ACD}|\boldsymbol{X}^{BCD}|\boldsymbol{X}^{ABCD}]$$

いうまでもなく $\boldsymbol{X}$ の行数は常に $IJKL$ であるが，列数はモデルによって変化する．ただし，飽和モデル (B.1) に関しては列数も $IJKL$ である．効果ベクトルとの対応関係は

$$\boldsymbol{\beta}^0\longleftrightarrow\boldsymbol{X}^0$$
$$\boldsymbol{\beta}^1\longleftrightarrow\boldsymbol{X}^A, \boldsymbol{X}^B, \boldsymbol{X}^C, \boldsymbol{X}^D$$
$$\boldsymbol{\beta}^2\longleftrightarrow\boldsymbol{X}^{AB}, \boldsymbol{X}^{AC}, \boldsymbol{X}^{AD}, \boldsymbol{X}^{BC}, \boldsymbol{X}^{BD}, \boldsymbol{X}^{CD}$$
$$\boldsymbol{\beta}^3\longleftrightarrow\boldsymbol{X}^{ABC}, \boldsymbol{X}^{ABD}, \boldsymbol{X}^{ACD}, \boldsymbol{X}^{BCD}$$
$$\boldsymbol{\beta}^4\longleftrightarrow\boldsymbol{X}^{ABCD}$$

で明白であろう．どの小行列も成分を $X_{sr}$ とすると，ダミーコーディングの下では $X_{sr}$ の値は 1 または 0 のいずれかである．そのなかでも，$\boldsymbol{X}^0$ は全部の要素が 1 になる例外的存在で，他の小行列は必ず 1, 0 が混在する．では，$\boldsymbol{X}^A, \boldsymbol{X}^{AB}$ を例に成分 $X_{sr}$ の値を決める規則を紹介しよう．

まず，

$$\boldsymbol{X}^A=[X_{sr}]$$

を考える．これは効果項 $\{u_i^A : i\in\underline{\mathscr{I}}\}$ に対応しているので，

$$s\in\mathscr{I}\times\mathscr{J}\times\mathscr{K}\times\mathscr{L},\qquad r\in\underline{\mathscr{I}}$$

の範囲で添字が変化する．ここで，規則によると，

$$\text{if}\quad s=\langle i^*, j, k, l\rangle\quad\text{then}\quad X_{sr}=0,$$
$$\text{else}$$

$$\text{if}\quad s\text{ の第 1 成分 } i \text{ が } r \text{ と等しい}\quad \text{then}\quad X_{sr}=1$$
$$\text{else}\qquad\qquad\qquad\qquad X_{sr}=0$$

と値が決められる．期待頻度あるいは効果項との対応を簡単にするため，添字 $r$ は上のように用いられたが，このままでは小行列の列を指定するには $i^*$ に相当するところが欠けていて不都合を生じる．値を決めた後は，$1, \cdots, (I-1)$ の通し番号の添字として調節する必要がある．次に

$$X^{AB}=[X_{sr}]$$

について調べてみよう．これは，交互作用 $\{u_{ij}^{AB} : ij \in \underline{\mathscr{I}} \times \underline{\mathscr{J}}\}$ に対応していて

$$s \in \underline{\mathscr{I}} \times \underline{\mathscr{J}} \times \underline{\mathscr{K}} \times \underline{\mathscr{L}}, \qquad r \in \underline{\mathscr{I}} \times \underline{\mathscr{J}}$$

である．値は

$$\text{if}\quad s=\langle i^*, j, k, l\rangle \quad \text{or} \quad s=\langle i, j^*, k, l\rangle \quad \text{or} \quad s=\langle i^*, j^*, k, l\rangle \quad \text{then} \quad X_{sr}=0,$$
$$\text{else}$$
$$\qquad \text{if}\quad s\text{ の成分 } (i, j) \text{ が } r \text{ と等しい}\quad \text{then}\quad X_{sr}=1$$
$$\qquad \text{else}\qquad\qquad\qquad X_{sr}=0$$

の規則によって割り当てられる．$r$ の番号の事後のつけかえに関しては先と同様である．

## 2. ANOVA コーディング

飽和モデルは

$$\log m_{ijkl}=u+u_i^A+u_j^B+u_k^C+u_l^D+u_{ij}^{AB}+u_{ik}^{AC}+u_{il}^{AD}+u_{jk}^{BC}+u_{jl}^{BD}+u_{kl}^{CD}$$
$$+u_{ijk}^{ABC}+u_{ijl}^{ABD}+u_{ikl}^{ACD}+u_{jkl}^{BCD}+u_{ijkl}^{ABCD}$$
$$(i \in \underline{\mathscr{I}},\ j \in \underline{\mathscr{J}},\ k \in \underline{\mathscr{K}},\ l \in \underline{\mathscr{L}}) \tag{B.2}$$

である．効果項は添字ごとにその和が 0 となるよう制約されている．

$$\sum u_i^A=\sum u_j^B=\sum u_k^C=\sum u_l^D=0$$
$$\sum_i u_{ij}^{AB}=\sum_i u_{ik}^{AC}=\sum_i u_{il}^{AD}=\cdots=\sum_l u_{il}^{AD}=\sum_l u_{jl}^{BD}=\sum_l u_{kl}^{CD}=0$$
$$\sum_i u_{ijk}^{ABC}=\sum_i u_{ijl}^{ABD}=\cdots=\sum_l u_{ikl}^{ACD}=\sum_l u_{jkl}^{BCD}=0$$
$$\sum_i u_{ijkl}^{ABCD}=\sum_j u_{ijkl}^{ABCD}=\sum_k u_{ijkl}^{ABCD}=\sum_l u_{ijkl}^{ABCD}=0$$

この制約の下では，冗長な項を他の項から導くことができる．たとえば

$$u_1^A=-(u_2^A+\cdots+u_1^A)$$

よって，変数ごとに冗長なカテゴリーを $(i^*, j^*, k^*, l^*)$ のように指定すれば，効果ベクトル $\boldsymbol{\beta}$ については，1.1(p. 201)で説明した構成原理がそのまま適用可能である．もちろん効果項は $u, u_i^A, \cdots, u_{ijkl}^{ABCD}$ におきかえる必要がある．

デザイン行列 $\boldsymbol{X}$ に関しては，小行列

$\boldsymbol{X}=$

$$[\boldsymbol{X}^0|\boldsymbol{X}^A|\boldsymbol{X}^B|\boldsymbol{X}^C|\boldsymbol{X}^D|\boldsymbol{X}^{AB}|\boldsymbol{X}^{AC}|\boldsymbol{X}^{AD}|\boldsymbol{X}^{BC}|\boldsymbol{X}^{BD}|\boldsymbol{X}^{CD}|\boldsymbol{X}^{ABC}|\boldsymbol{X}^{ABD}|\boldsymbol{X}^{ACD}|\boldsymbol{X}^{BCD}|\boldsymbol{X}^{ABCD}]$$

の構成が一部異なる．ANOVA コーディングの下では，$\boldsymbol{X}^0$ を除く小行列は

$$-1,\quad 0,\quad +1$$

のいずれかの値を取る．$\boldsymbol{X}^A, \boldsymbol{X}^{AB}$ を例に成分 $X_{sr}$ の値を決める規則を紹介しよう．

まず,

$$X^{A}=[X_{sr}] \qquad (s\in \underline{\mathscr{I}}\times\underline{\mathscr{J}}\times\underline{\mathscr{K}}\times\underline{\mathscr{L}},\ r\in\underline{\mathscr{I}})$$

を考える．規則によると，

if　$s=\langle i^*,j,k,l\rangle$　then　$X_{sr}=-1$,

　else

　　if　$s$ の第 1 成分 $i$ が $r$ と等しい　then　$X_{sr}=1$

　　　else　　$X_{sr}=0$

と値が決められる．続いて

$$X^{AB}=[X_{sr}] \qquad (s\in \underline{\mathscr{I}}\times\underline{\mathscr{J}}\times\underline{\mathscr{K}}\times\underline{\mathscr{L}},\ r\in\underline{\mathscr{I}}\times\underline{\mathscr{J}})$$

について調べてみよう．値は

if　$s=\langle i^*,j,k,l\rangle$　or　$s=\langle i,j^*,k,l\rangle$　or　$s=\langle i^*,j^*,k,l\rangle$　then　$X_{sr}=-1$,

　else

　　if　$s$ の成分 $(i,j)$ が $r$ と等しい　then　$X_{sr}=1$

　　　else　　$X_{sr}=0$

なる規則に従って決められる．$r$ の事後の調整に関してもダミーコーディングの場合と同じことがいえる．

## C. コンピュータプログラムの紹介

対数線型モデルと潜在クラスモデルを実際のデータ分析に応用するにあたって簡単に利用できるソフトウェアを紹介する．

### 1. 対数線型モデル

① SPSS-X (Statistical Package for the Social Sciences) の LOGLINEAR は一般のユーザに親しみやすい操作性とデザイン行列の指定が柔軟性に富んでいる (cf. 三宅と山本, 1988).

② SAS (Statistical Analysis System) の CATMOD は特に Logit 分析に適しているが，使用に際してはやや高度な知識が要求される．

③ ECTA (Everyman's Contingency Table Analyzer) は Fay と Goodman (1975) によって FORTRAN で書かれたもので，パーソナルコンピュータに移植可能である．効果は全平均 $u$ からの偏りによって定義されている．

④ GLIM (Generalized Linear Interactive Modeling) は Baker と Nelder (1978) が一般線型モデル解析用に開発した汎用パッケージで，対数線型モデルを回帰モデル等他の線型モデルとの関連を把握するには適しているかもしれない．効果は全基準 $u$ からの偏りで定義されるが，基準セルは各変数の第 1 水準に固定されていて柔軟性に欠ける．

### 2. 潜在クラスモデル

Clogg (1977) の開発した MLLSA (Multivariate Latent Structure Analysis) が広く普及している．FORTRAN で書かれているので必要に応じて容易に改良できる．